Eternal Vigilance is the Price of Performance

Robert P. Colwell

Konrad Lai

Intel Corp.

June, 1995

Pre-Silicon Words To Live By

Maxim 1:

If it wasn't tested, it doesn't work.

Maxim 2:

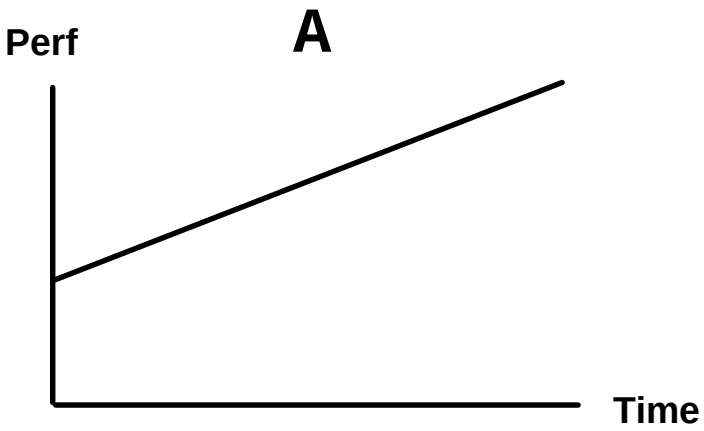
It doesn't work until it works at speed.

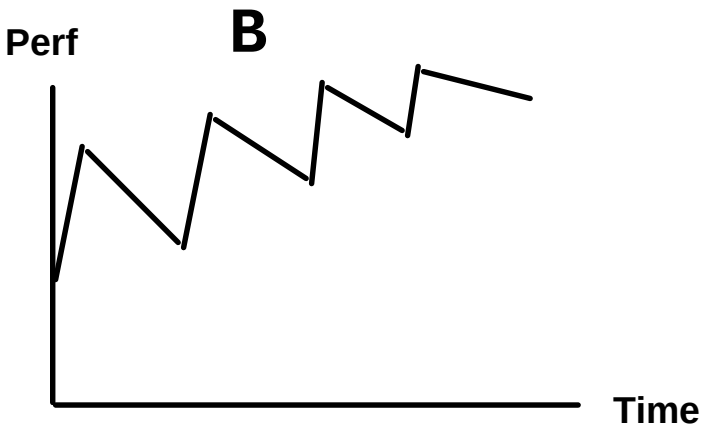
Pre-silicon validation must include both function and performance

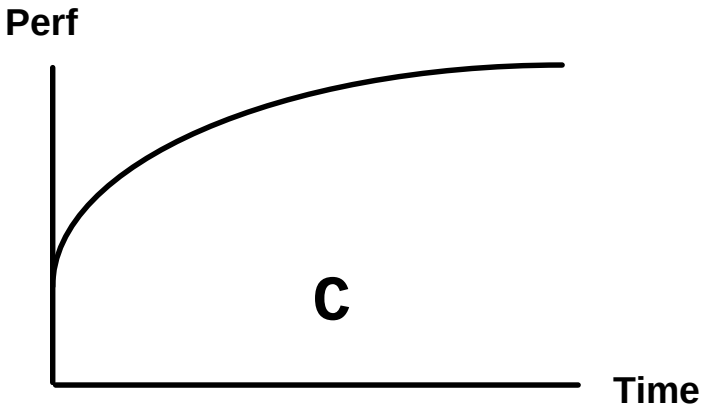
intel

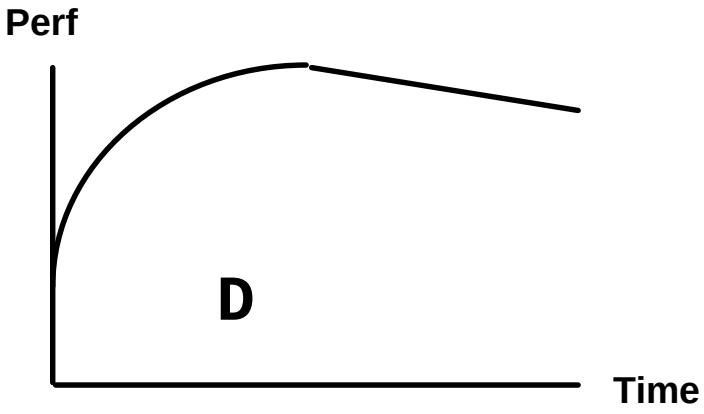
You're starting a new CPU design...

...what performance-vs-time would you expect?









Root cause

- Complex machines exhibit complex behavior
- Aggregate of thousands of local design decisions
  - Each choice *looks* innocuous to designer
  - Sometimes even looks innocuous to architect
  - But cause a surprising effect on performance

Surprises are almost always negative!

What to do? Perf analysis

- Regressions
  - running, analyzing, improving
  - integer, FP, desktop, MP
- Projections
  - tools/methods
  - calibration: when to trust

Tools of the trade

- DFA
- P6PS
- RTL
- Analytic projections (MP)
- All tools have
  - Jumpstart capability
  - Visualization facilities
  - Automatic regressions
  - Automatic perf error checkers

DFA (DataFlow Analyzer)

- Tracks true data dependencies
  - All else is superimposed on it (HW limits, implem. artifacts)
- Trace-based
- Runs at 1K instrs/sec
- Highly tuneable & reconfigurable
- Especially useful in early uArch explorations
- Not a simulator; doesn't actually execute code
- Watch out for artifacts and inaccurate results

Functional Simulator

- Quite accurate, executes code
  - Results can be checked
- 600 instrs/sec
- Models implementation details
  - Fidelity high
  - Cost of modelling also high
  - Must track chip development
  - Can study speculative effects

RTL

- Dog Slow: 3 Hz
  - Limits size of studies
  - Cold-start problems (cache, TLB, BTB state)
- Completely accurate (if you include sys details too)
- Changes weekly in random ways
  - Functional bugs can kill perf analysis
- Can't export model to get help from elsewhere
- Can't explore design space
  - Don't even *have* RTL until design is very far along
  - Options to fix perf bugs limited when found late in game
- Hard to create perf benchmarks for RTL
  - Hard to make big enough for perf effect, small enough to run

Visualization

```
#32   :D003A281:MOVw E:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
# Old Time:                    200-------210-------220-------230-------240-------250-------260-------270
# X-Bus   ( 200  283)            (   C 8 8 8 8 X C B A A A A X B B B X C C X       E E E E X         D D D D X   )
# New Time:                    280-------290-------300-------310-------320-------330-------340-------350
#   54:00182858:LDw(M,:----:NOPRB:MM: +  DdARVvALMB-------M-X--------------------W-R-----C .   +   .   +
#33   :D003A284:ORw Ib:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   55:--------:OR(GRS:----:-----:--: +  DdARVv.....................................EW,R  +   .   +   .   +
#34   :D003A286:MOVw R:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   56:--------:STDATA:----:-----:--: +   DdARVv.....................................DW,R +   .   +   .   +
#   57:00182858:STADDR:----:-----:--: +   DdARVvASSW,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,R +   .   +   .   +
#35   :D003A28D:MOVw I:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   58:--------:STDATA:----:-----:--: +   DdARVvDW,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,R +   .   +   .   +
#   59:00182880:STADDR:----:-----:--: +   DdARVvASSW,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,R+   .   +   .   +
#36   :D003A28E:POPw B:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   60:00881BC0:LDw((T:----:NOPRB:H : +   .DdARVvAL1W,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,R+   .   +   .   +
#   61:--------:ADD((S:----:-----:--: +   .DdARVvEW,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,R+   .   +   .   +
#37   :D003A28F:POPw B:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   62:00881BC4:LDw((T:----:NOPRB:H : +   . DdARVvAL1W,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,R   .   +   .   +
#   63:--------:ADD((S:----:-----:--: +   . DdARVvEW,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,R   .   +   .   +
#    :D003A2A0:------:STRM:1   :MM: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#------:D003A290:------:NEWL:-----:H : +FIQ .   +   .   +   .   +   .   +   .   +   .   +   .   +
#    :--------:------:----:-----:--: +   . B +   .   +   .   +   .   +   .   +   .   +   .   +
```

Visualization cont.

```
#70   :D00C9A80:XCHGw :----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   139:00147ADC:LDw(M,:----:NOPRB:H : +   .   + DdARVvAL1W,R.   +   .   +   .   +   .   +   .   +
#   140:--------:MOV((T:----:-----:--: +   .   + DdARVv...EW,R   +   .   +   .   +   .   +   .   +
#   141:--------:STDATA:----:-----:--: +   .   + DdARVv....DW,R   +   .   +   .   +   .   +   .   +
#   142:00147ADC:STADDR:----:-----:--: +   .   + Dd~ARVvASSW,,R   +   .   +   .   +   .   +   .   +
#71   :D00C9A86:CMPw R:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   143:00147AE0:LDw(M,:----:NOPRB:H : +   .   +   DdARVvAL1W,,R   +   .   +   .   +   .   +   .   +
#   144:--------:SUB((T:----:-----:--: +   .   +   DdARVv...EW,,R   +   .   +   .   +   .   +   .   +
#72   :D00C9A88:JNBE I:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   145:--------:JMP(FL:FNM+:-----:--: +   .   +   Dd~ARVv...JW,R   +   .   +   .   +   .   +   .   +
#73   :D00C9A89:   STI:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   146:--------:MOV(():----:-----:--: +   .   +   DdARVvEW,,,,R   +   .   +   .   +   .   +   .   +
#   147:--------:MOV(FL:----:-----:--: +   .   +   DdARVv...EW,,R +   .   +   .   +   .   +   .   +
#   148:--------:MOV(T1:----:-----:--: +   .   +   Dd~ARVv...EW,R +   .   +   .   +   .   +   .   +
#   149:--------:OR(T0,:----:-----:--: +   .   +   Dd-ARVvEW,,,,R +   .   +   .   +   .   +   .   +
#   150:--------:MOV(():----:-----:--: +   .   +   .DdARVv.EW,,,,R+   .   +   .   +   .   +   .   +
#   151:--------:MOV(T0:----:-----:--: +   .   +   .Dd~ARVv..EW,,R+   .   +   .   +   .   +   .   +
#   152:--------:SHFT((:----:-----:--: +   .   +   .Dd-ARVv...EW,R+   .   +   .   +   .   +   .   +
#   153:--------:AND(T1:----:-----:--: +   .   +   .Dd-ARVv....EW,R   .   +   .   +   .   +   .   +
#   154:--------:ADD(T0:----:-----:--: +   .   +   . Dd~ARVv..EW,,R   .   +   .   +   .   +   .   +
# *142:00147ADC:MOB_ST:----:-----:H : +   .   +   .   +   mH1 +   .   +   .   +   .   +   .   +
#   155:--------:SYNC():----:-----:--: +   .   +   . Dd-~!!!!!!!!!ARVvYW,R +   .   +   .   +   .   +
#74   :D00C9A8A:  RETn:----:-----:--: +   .   +   .   +   .   +   .   +   .   +   .   +   .   +
#   156:00881C08:LDw((T:----:NOPRB:H : +   .   +   . Dd----------ARVvAL1W,R+   .   +   .   +   .   +
#   157:--------:ADD((S:----:-----:--: +   .   +   . Dd----------ARVv.EW,,R+   .   +   .   +   .   +
#   158:--------:JMP(T0:REM+:-----:--: +   .   +   . Dd----------~ARVv..JW,R   .   +   .   +   .   +
#------:D0027514:------:BRSP:-----:H : +   .   +   .   +FIQ .   +   .   +   .   +   .   +   .   +
#    :--------:------:----:-----:--: +   .   +   .   + B .   +   .   +   .   +   .   +   .   +
```

Pitfalls

- OOO/Spec-ex with trace-based analysis
  - Real workloads
    - Real apps (database), real OS kernel
  - Include I/O, TLBs, memory, bus overhead
  - Heuristics fill in for mis-speculation
- Include the bus & chipset
- Can't do performance until functionality is there
  - So performance work is chronically behind schedule

Pitfalls, cont.

- Crucial! Must calibrate tools to each other
  - RTL results are trustworthy
    - Functionally checked at runtime, after all
    - But can't run big programs through
  - When all tools agree on small programs
    - Confidence in "big program" tools increases
    - Not otherwise checkable
  - Tools disagree? You're about to learn something
  - Correlation is initially human-expert intensive
  - But must be automated to get enough trials done

Summary

- All tools have their place
  - RTL: gotta do it, but can't be *all* you do
  - DFA: light-weight exploration tool extremely valuable
    - But only in hands of an expert
    - Even then, don't always believe everything it says
  - Fnl Sim: useful for confirming implementation performance
  - Analytic: useful for system-level (esp. MP)
- Constant vigilance needed to keep design effort on track
  - Hundreds of DE's are trying to push it off
  - Judgment required: when to fight vs. when to accept perf loss

Machines are too complex to take performance for granted any more